*Qualitative and Quantitative Methods in Libraries (QQML) 13,1:99-124, 2024*

# Modelling the Growth of Indo-US Collaborated Literature in the Field of Bioinformatics

**Pooja Pooja[1], Prof. Meera[2]**
[1,2]University of Delhi

**ABSTRACT :** This study highlights the growth of literature produced as a result of Indo-US collaboration in the bioinformatics domain covering the period 1999-2021. A total of 719 publications have been produced through Indo-US collaboration in bioinformatics, the data for which were downloaded using Scopus and subjected to analysis using MS Excel software. To examine the growth pattern and trend in publications, various growth models were applied to the literature. Overall, an increasing trend has been observed, and the declining RGR tends to reflect the slow but steady increase in the number of articles published, as evidenced by an increase in DT. The growth of literature produced as a result of Indo-US partnership in the area of bioinformatics follows the Linear Growth Model. When the different growth models using curve fit values based on $R^2$ have been applied, the Power Growth Curve is found to be the best match, followed by the Exponential Growth Curve, Linear Growth Curve, and Logarithmic Growth Curve.

**KEYWORDS**: Annual Growth Rate, Relative Growth Rate, Doubling Time, Linear Growth Model, Power Growth Model, Exponential Growth Model, Logarithmic Growth Model

## INTRODUCTION

According to experts, one of the most striking aspects of modern research is the incredible growth in scientific findings and knowledge. This has resulted in an unparalleled build-up of data, which has been a major source of concern for experts in various fields as well as academicians and research scholars. As a result, there is a requirement to investigate the evolution of scientific knowledge and literature in a

Received: 22.4.2023 Accepted: 30.1.2024 ISSN 2241-1925

particular field. The term ‘growth’ denotes a rise in real size, which further signifies an overall change in a state. When particularly observing the field of science and technology, ‘growth’ can refer to an increase in the number of organisations, researchers, or publications, among other things, while if we talk in terms of literature, then it refers to a change in the number of publications over a given time frame. Fitting growth models and curves to data on rapidly increasing literature and determining the best fit to describe the expansion of literature is an essential element of growth research. Modern biological research has made access to genetic information and synthesising it for the discovery of new knowledge a fundamental focus. The use of modern computational algorithms is required for mining genetic data. For this reason, in order to tackle the genomic era’s new challenges, it is critical for the next generation of biologists to get acquainted with an area of research concerned with the careful preservation, organisation, and indexing of information. Bioinformatics is a branch of science that uses information from biological experiments to analyse, comprehend, and model the entire life process as an information-processing phenomenon over energy.

For numerous years, scientific partnerships between India and the USA have been fruitful thanks to a range of bilateral agreements. The collaborative effort in bioinformatics research signifies India’s advancement in this field with other top-most nations working on similar projects. There are several examples of India-US cooperation focusing on epigenetics, fundamental biology, and ophthalmic illnesses, as well as visual problems. Like the other fields of science, the literature on bioinformatics has also expanded over the years. The published literature or research publications serve as a gauge for measuring knowledge in a particular domain, and a study of publication growth rates might give some vital information regarding the same. So, the pattern of growth of literature in bioinformatics that has been produced as a result of collaboration between India and the USA has been analysed by applying various growth models and scientific indicators of growth.

## LITERATURE REVIEW

El-Shorbagy and El-Refaey (2022) investigated the impact of preventative measures on mathematical development and the COVID-19 epidemic’s propagation using different growth models like exponential, geometric and logarithmic. They discovered that in China, the stringent preventative actions of isolating, forbidding all meetings, and imposing a complete curfew caused the epidemic’s spread to shift from an exponential to a growth pattern more like a logarithmic model while in the Kingdom of Saudi Arabia and the USA COVID-19 expanded significantly during the early phase of the dissemination without stringent preventative measures, which was typical of most epidemics. However, adopting reasonable preventive measures and people’s dedication to putting them into practice had slowed down the epidemic’s exponential development rate, as was the case in the Kingdom of Saudi Arabia while failing to do so had sped up the epidemic’s growth pace, as was the case in the USA.

Jozi and Nourmohammadi (2022) analysed the number of patents, scholarly papers, and patent applications for pathology and forensic medicine based on *Scopus* from 2011- 2020. The findings indicated a lack of academic-industry collaboration in pathology and forensic medicine, even though North America is indeed the busiest location in this area. A positive exponential relationship was observed between the average increase in patents based on publications, the number of articles cited in patents, patent citations, and the typical number of patents issued to an organisation in its publications. According to the linear model, there seemed to be a negative and inverse association between articles and citations. The article and citation data set were well-suited for the polynomial, linear and exponential growth models.

Singh and Verma (2021) used Egghe and Rao’s growth rate functions and discovered the best-suited growth model in the field of food science and technology in India. According to their findings, the growth of the food science and technology literature was feasible in terms of both publications and citations.

Elangovan (2020) applied the goodness-of-fit test using linear and non-linear growth models to medical literature produced by the faculty of AIIMS and tested the growth using the curve-fitting approach. They discovered that in comparison to the other growth models, the exponential growth model best matched the literature. The faculty size and department did not have a notable impact on the efficiency of AIIMS, according to the t statistics, which demonstrated that they were significant independent variables when compared to the department, although they do not approach the statistical significance value of p equal to 0.344.

Neelamma and Anandhalli (2020) used data from the *Web of Science* Database to highlight the expansion of crystallography literature from 1989 to 2013. Different growth models were applied, and it was found that the relative growth rate of crystallography publications was falling, while the doubling time was increasing. The expansion of literature on the subject of crystallography did not match the exponential or logistic growth models; however, it closely matched the polynomial, power, and linear models, and a steady pattern of growth was observed.

Nayak and Bankapur (2017) sought to trace the rise of agricultural literature at the global and national levels using several growth models, as well as a comparison of the growth and dynamics of the top 10 nations in the area from 1930 to 2016. Their research also focused on relative growth rate, doubling time, regression, as well as other growth parameters, to determine whether agricultural literature matches exponential, linear, logistic, or power growth models. They found that the literature followed the linear and exponential growth models and there has been a constant tendency toward higher expansion of agricultural literature.

Hadagali and Anandhalli (2015) analysed the evolution of neurology literature for fifty years (1961-2010), utilising data from *Science Direct* and different growth patterns along with relative growth rate and doubling time to determine which growth model better suited the neurology literature. It was found that the literature did not match either of the linear or logistic growth models. However, it closely resembled the exponential growth model as a steady growth was observed over the years.

Sangam, Madalli, and Arali (2015) aimed to investigate the growth pattern of the Indian and global genetics literature (1993-2012). Their study concluded that the world's genetics literature followed the logarithmic as well as the linear growth models. On the other hand, the exponential and logistic models fitted well for India.

**Sharma, Gupta, and Kumar (2002)** used growth models to examine global literature in physics, chemistry, and electrical and electronic engineering studies from 1907 to 1994. The growth models were categorised and defined using Egghe and Rao's proposed two growth-rate functions, and the literature was further analysed using their methodology. The outcomes of the analysis were not entirely consistent with the conclusions predicted by Egghe and Rao's technique.

## METHODOLOGY

The data for the study were collected using the *Scopus* database, and 'bioinformatics' was used as the keyword to get the search results. Filters were applied to refine and limit our search to different countries and years, i.e., the search is limited to the USA and India. Individually, the USA produced 43,499 publications in bioinformatics and India produced 7,134 publications. The objective of the current study is to analyse the growth pattern and apply growth models on Indo-US collaborated literature of bioinformatics, so only the collaborated number of papers from both countries, i.e., 719 publications, were studied for detailed analysis. The data were extracted in CSV format from *Scopus* and analysed using\ MS Excel software.

## OBJECTIVES

1. To analyse the chronological distribution of research publications produced as a result of Indo-US collaboration.

2. To observe the growth pattern and find AGR (annual growth rate), RGR (relative growth rate), and DT (doubling time) of the Indo-US collaborated literature of bioinformatics.
3. To apply various growth models to the literature of the bioinformatics domain produced in collaboration from India and the USA.

## HYPOTHESES

1. **Hypothesis I** states that "*There is an overall increasing trend in literature produced as a result of collaboration from India and the USA in the bioinformatics domain over the period of time*".
2. **Hypothesis II** states that "*There is a linear growth of literature in bioinformatics produced as a result of Indo-US collaboration*".

## GROWTH PATTERN AND GROWTH MODELS

Various growth strategies are used to examine literature's development, growth, and dynamism, in the bioinformatics subject area, and numerous growth models and scientific growth indicators are used to depict how the volume of literature changed over the years.

i. **Annual Growth Rate (AGR):** It is used to compute the rate of increase in the number of publications every year. The yearly growth rate of publications or the annual growth rate is derived using Kumar and Kaliyaperumal's (2015) methodology, which is listed below:

$$\textbf{AGR} = \frac{End\ value - First\ value}{First\ value} \times \textbf{100}$$

ii. **Relative Growth Rate (RGR):** The surge in the no. of papers per unit of time is referred to as relative growth rate, i.e., RGR. When it comes to scientific literature production, it is also known as the exponential or continuous growth rate. Moreover, this technique may also be used to compute the average relative growth rate of papers during a certain period. The RGR and DT models, which were established by Mahapatra in 1985, were used to calculate the rate of growth of all publications (Mahapatra, 1985). The formula for RGR is:

**RGR** = $\frac{W2 - W1}{T2 - T1}$

RGR signifies the rate of growth throughout the span;

W1 signifies the natural log ($Log_e$ ) of the initial number of publications;

W2 signifies the natural log ($Log_e$) of the final number of publications;

T1 signifies the initial time unit;

T2 signifies the final time unit (Shilpa and Padmamma, 2020).

**iii.** **Doubling Time (DT):** The amount of time needed for a variable to double in size or amount, or we can say the time when the number of articles in a specific category doubles, is known as the doubling time, which is inversely proportional to the relative growth rate, with RGR being constant. The amount grows at an exponential rate, with a constant doubling time that can be determined immediately from the rate of growth. As a result, the doubling time is computed as:

**DT = 0.693 / R** where,

DT denotes the doubling time;

R denotes the relative growth rate of the publications.

**iv.** **Linear Growth Model:** Shorter periods of growth can be approximated using the linear growth model. The progression rate in this example is linear, which means that the constant growth per unit of time is represented as:

**y = mx + b**

where m denotes the slope and b is the intercept (Nayak & Bankapur, 2017, p.102).

v. **Exponential Growth Model:** Whenever the rate of expansion or growth is proportional to the rise of papers per unit of time, an exponential model matches the dataset. The least squares fit via points are calculated using the exponential growth pattern equation that can be represented as follows:

$y = ce^{bx}$

where c and b both are the constants and e is said to be the base of the natural log

(Nayak & Bankapur, 2017, p. 102).

vi. **Power Growth Model:** A functional relation among two quantities changes with the magnitude or power of the other, as per Rao (2010). The least-squares fit across points is calculated using the following equation of the Power Growth model: $y = cx^b$ where both the c and b are constants (Nayak & Bankapur, 2017, p.102).

vii. **Logarithmic Growth Model:** The equation for the logarithmic growth model can be expressed as $y = c \ln x + b$ where c and b are both constant values and *ln* denotes the natural log function (Nayak & Bankapur, 2017, p.102).

## DATA ANALYSIS AND INTERPRETATION

### 1. CHRONOLOGICAL DISTRIBUTION OF PUBLICATIONS AND GROWTH RATE

**Table 1. Chronological Distribution of Collaborated Publications between India and USA & Growth Rate**

| Year | No. of publications | % of publications | Cumulative publications | AGR | W1 (ln of initial value) | W2 (ln of final value) | RGR = (W2-W1) | DT = (0.693/R) |
|---|---|---|---|---|---|---|---|---|
| **1999** | 1 | 0.14 | 1 | | | 0.00 | | |
| **2002** | 1 | 0.14 | 2 | 0.00 | 0.00 | 0.69 | 0.69 | 1.00 |
| **2003** | 6 | 0.83 | 8 | 500.00 | 0.69 | 2.08 | 1.39 | 0.50 |
| **2004** | 7 | 0.97 | 15 | 16.67 | 2.08 | 2.71 | 0.63 | 1.10 |
| **2005** | 12 | 1.67 | 27 | 71.43 | 2.71 | 3.30 | 0.59 | 1.18 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **2006** | 3 | 0.42 | 30 | -75.00 | 3.30 | 3.40 | 0.11 | 6.58 |
| **2007** | 10 | 1.39 | 40 | 233.33 | 3.40 | 3.69 | 0.29 | 2.41 |
| **2008** | 25 | 3.48 | 65 | 150.00 | 3.69 | 4.17 | 0.49 | 1.43 |
| **2009** | 19 | 2.64 | 84 | -24.00 | 4.17 | 4.43 | 0.26 | 2.70 |
| **2010** | 26 | 3.62 | 110 | 36.84 | 4.43 | 4.70 | 0.27 | 2.57 |
| **2011** | 21 | 2.92 | 131 | -19.23 | 4.70 | 4.88 | 0.17 | 3.97 |
| **2012** | 25 | 3.48 | 156 | 19.05 | 4.88 | 5.05 | 0.17 | 3.97 |
| **2013** | 44 | 6.12 | 200 | 76.00 | 5.05 | 5.30 | 0.25 | 2.79 |
| **2014** | 53 | 7.37 | 253 | 20.45 | 5.30 | 5.53 | 0.24 | 2.95 |
| **2015** | 51 | 7.09 | 304 | -3.77 | 5.53 | 5.72 | 0.18 | 3.77 |
| **2016** | 42 | 5.84 | 346 | -17.65 | 5.72 | 5.85 | 0.13 | 5.36 |
| **2017** | 55 | 7.65 | 401 | 30.95 | 5.85 | 5.99 | 0.15 | 4.70 |
| **2018** | 62 | 8.62 | 463 | 12.73 | 5.99 | 6.14 | 0.14 | 4.82 |
| **2019** | 76 | 10.57 | 539 | 22.58 | 6.14 | 6.29 | 0.15 | 4.56 |
| **2020** | 89 | 12.38 | 628 | 17.11 | 6.29 | 6.44 | 0.15 | 4.53 |
| **2021** | 91 | 12.66 | 719 | 2.25 | 6.44 | 6.58 | 0.14 | 5.12 |
| **Grand Total** | **719** | **100.00** | **Average=** | **53.49** | | | **0.33** | **3.30** |

Table 1 and Figure 1a show the chronological output of Indo-US partnership in the bioinformatics sector from 1999 to 2021. A total of 719 publications was reported, with an average of 34 publications per year. The years with the most and lowest number of publications are 2021 (91 articles, 12.66%) and 1999 and 2002 (1 paper each, 0.14%). The average AGR, RGR, and DT values, respectively, are 53.49, 0.33, and 3.30.

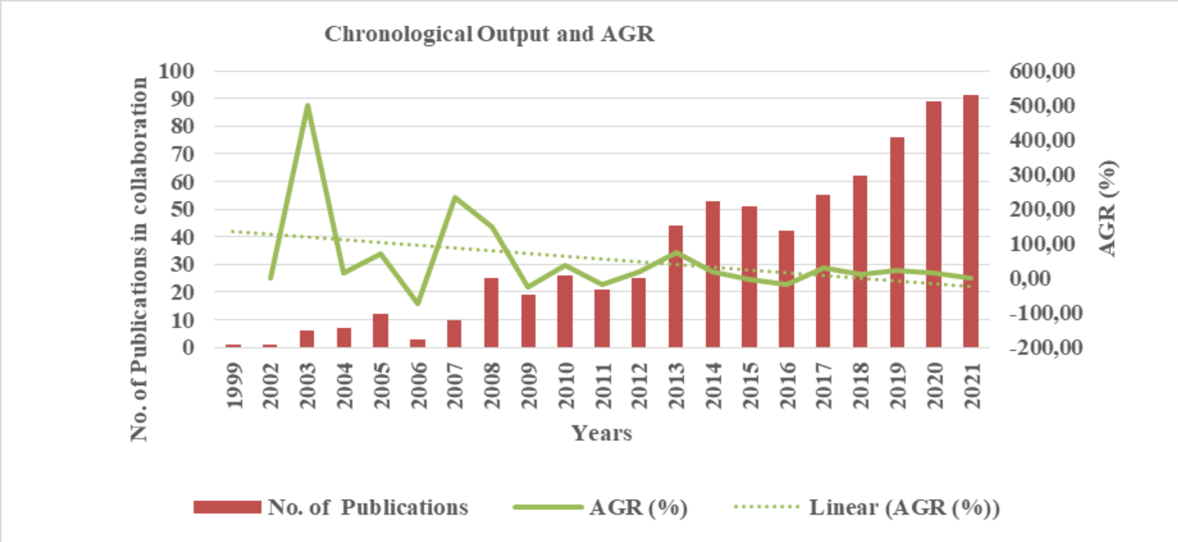


**Figure 1(a). Chronological Publication Output and Growth Rate between India and USA**

A fluctuating growth in publications produced in collaboration from both nations can be observed from the data. The average annual growth rate (AGR) is 53.49, indicating an overall upward trend in the growth of publications in bioinformatics. Furthermore, the declining RGR tends to reflect the slow but steady increase in the number of articles published, as evidenced by an increase in DT, which is significantly greater than RGR during the course of the study.

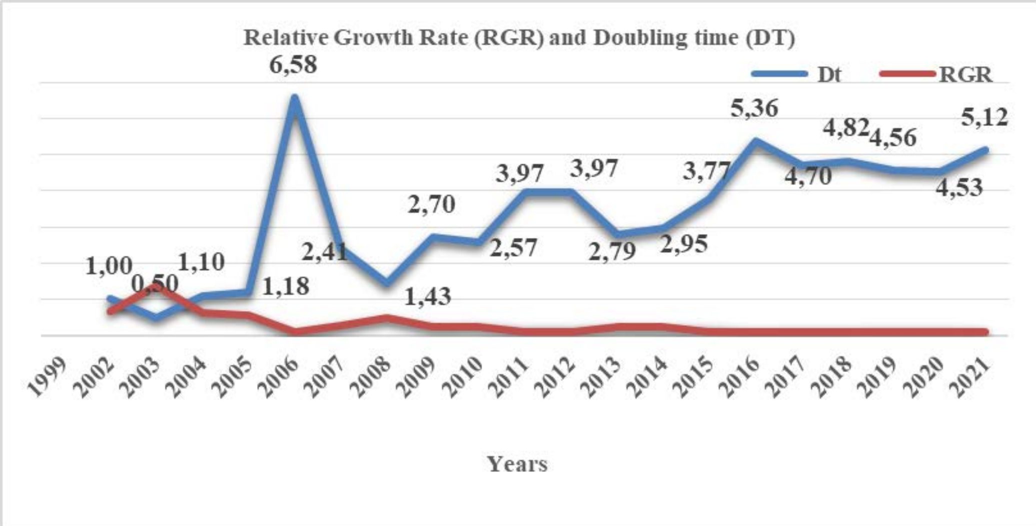


**Figure 1(b). Relative Growth Rate and Doubling Time of Indo-US Collaborated Publications**

In a study conducted by Baskaran (2021) covering bioinformatics research during 2010- 2019, he concluded that RGR decreased from 0.75 in 2011 to 0.20 in 2018, and the doubling time of the publications gradually increased from 0.92 in 2011 to 3.47 in 2018. Also, Baskaran (2016) conducted a scientometric study of bioinformatics literature from 1999-2013 and concluded that RGR and DT exhibit a fluctuating trend throughout the study.

Manoharan et al. (2011) analysed the bioinformatics literature for a period of 11 years, which yielded similar results, though not consistently, in that research output in bioinformatics was rising.

## 1.1 MANN-KENDALL TEST (M-K TEST)

M-K test has been employed to statistically prove if there is an overall increasing or decreasing trend in literature produced as a result of Indo-US collaboration in bioinformatics. Any positive value of the test statistic Z indicates a growing trend in the data under consideration, whereas any negative value shows a decreasing trend. This test is non-parametric in nature.

First of all the n time series values ($X_1$, $X_2$….$X_n$.) are replaced by their relative ranks ($R_1$, $R_2$….$R_n$)

**Table 1.1 Calculation of S for M-K Test**

| Year | No. of Publications (X) | Rank (R) | S |
|---|---|---|---|
| 1999 | 1 | 1 | 19 |
| 2002 | 1 | 1 | 19 |
| 2003 | 6 | 4 | 16 |
| 2004 | 7 | 5 | 15 |
| 2005 | 12 | 7 | 12 |
| 2006 | 3 | 3 | 15 |
| 2007 | 10 | 6 | 14 |
| 2008 | 25 | 10 | 8 |
| 2009 | 19 | 8 | 12 |
| 2010 | 26 | 12 | 7 |
| 2011 | 21 | 9 | 10 |
| 2012 | 25 | 10 | 9 |
| 2013 | 44 | 14 | 6 |
| 2014 | 53 | 16 | 3 |
| 2015 | 51 | 15 | 4 |
| 2016 | 42 | 13 | 5 |
| 2017 | 55 | 17 | 4 |
| 2018 | 62 | 18 | 3 |
| 2019 | 76 | 19 | 2 |
| 2020 | 89 | 20 | 1 |
| 2021 | 91 | 21 | 0 |
| | | | **S = 184** |

In the above Table 1.1, the value of S has been calculated using the formula:

**S =** $\boldsymbol{\Sigma_{i=1}^{n-1} \Sigma_{j=k+1}^{n} sign\left(x_j - x_k\right)}$

**Sign (x) = {1 if x > 0, 0 if x = 0, -1 if x < 0}**

**S = 184**, where S is the no. of positive differences minus the no. of negative differences.

**n = 21** (No. of observations)

After finding the value of S, the variance of S has been calculated using the formula:

**Variance (S) = 1/18 [n(n-1) (2n+5) - $\Sigma_{p-1}^{g} t_p (t_p - 1)(2t_p + 5)$]**

Where g is the no. of tied groups,

$t_p$ is the no. of observations in the $p^{th}$ group

putting values of $t_1 = 1$, $t_2 = 25$

**Variance (S) = 1094.667**

Now, substituting the values in the following formula, $z_{Mk}$ is calculated as:

$z_{Mk} = (S - 1) \div \sqrt{var(S)}$ **if S > 0**

**= 0 if S = 0**

**= (S+1) $\div \sqrt{var(S)}$ if S < 0**

$z_{Mk}$ **= 5.531083**

Calculating z critical at 5% significance level, i.e., **Alpha (α) = 5%**

**z critical value is = 1.959964**

**(z-value at 99% confidence level = 2.58)**

From the above results, it is clear that z statistics value is positive and also greater than the z critical value; hence, a positive or increasing trend occurs in the dataset and the trend is significant at 99% confidence level.

**Hypothesis I** stating that "*There is an overall increasing trend in literature produced as a result of collaboration from India and the USA in the bioinformatics domain over the period of time*" stands supported.

## 2. APPLICATION OF GROWTH MODELS USING CHI-SQUARE TEST AND CURVE FIT VALUES

### 2.1 APPLICATION OF GROWTH MODELS USING CHI-SQUARE TEST

**Table 2.1 Linear Growth Model using Chi-Square**

| Sr. no. | Year | Observed no. of papers | Cumulative frequency Y (f) | Expected no. of papers (p) = mx+b | (f- p) | $(f-p)^2$ | $(f-p)^2/p$ |
|---|---|---|---|---|---|---|---|

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 1999 | 1 | 1 | -121.29 | 122.29 | 14954.35 | -123.296 |
| 2 | 2002 | 1 | 2 | -87.63 | 89.63 | 8032.82 | -91.6716 |
| 3 | 2003 | 6 | 8 | -53.96 | 61.96 | 3839.54 | -71.15 |
| 4 | 2004 | 7 | 15 | -20.30 | 35.30 | 1246.23 | -61.3847 |
| 5 | 2005 | 12 | 27 | 13.36 | 13.64 | 186.05 | 13.92587 |
| 6 | 2006 | 3 | 30 | 47.02 | -17.02 | 289.75 | 6.161977 |
| 7 | 2007 | 10 | 40 | 80.68 | -40.68 | 1655.19 | 20.51445 |
| 8 | 2008 | 25 | 65 | 114.35 | -49.35 | 2435.03 | 21.29526 |
| 9 | 2009 | 19 | 84 | 148.01 | -64.01 | 4097.02 | 27.6811 |
| 10 | 2010 | 26 | 110 | 181.67 | -71.67 | 5136.59 | 28.27428 |
| 11 | 2011 | 21 | 131 | 215.33 | -84.33 | 7111.89 | 33.02754 |
| 12 | 2012 | 25 | 156 | 248.99 | -92.99 | 8647.88 | 34.73129 |
| 13 | 2013 | 44 | 200 | 282.66 | -82.66 | 6832.01 | 24.17077 |
| 14 | 2014 | 53 | 253 | 316.32 | -63.32 | 4009.17 | 12.67449 |
| 15 | 2015 | 51 | 304 | 349.98 | -45.98 | 2114.16 | 6.040803 |
| 16 | 2016 | 42 | 346 | 383.64 | -37.64 | 1416.92 | 3.69334 |
| 17 | 2017 | 55 | 401 | 417.30 | -16.30 | 265.82 | 0.636995 |
| 18 | 2018 | 62 | 463 | 450.97 | 12.03 | 144.82 | 0.321127 |
| 19 | 2019 | 76 | 539 | 484.63 | 54.37 | 2956.31 | 6.100172 |
| 20 | 2020 | 89 | 628 | 518.29 | 109.71 | 12036.28 | 23.22307 |
| 21 | 2021 | 91 | 719 | 551.95 | 167.05 | 27905.03 | 50.55699 |
| | **Grand Total** | **719** | **4522** | | | | **-34.473** |

The above table depicts the application of the linear growth model to the dataset of Indo-US collaborated papers in bioinformatics during 1999 to 2021. The chi-square goodness of fit test has been applied to the above dataset.

**Linear growth model equation: y = mx+b**

Linear growth equation (y = 33.662x - 154.95)

**$R^2$ = 0.8833**

The linear growth model equation, i.e., y = 33.662x - 154.95 and $R^2$ = 0.8833, is calculated using Excel statistical operations.

m = 33.662

b = -154.95

Calculated chi-square value **= -34.473**

**df = 20**

Theoretical chi-square value for df (20) at 0.05 level of significance **= 31.41**

The calculated chi-square value, i.e., -34.473, is less than the theoretical value of chi-square at df = 20, i.e., 31.41; *hence, the growth of literature produced as a result of Indo-US partnership in the area of bioinformatics follows the linear growth model.*

**The hypothesis** stating that *"There is a linear growth of literature in bioinformatics produced as a result of Indo-US collaboration"* stands supported.

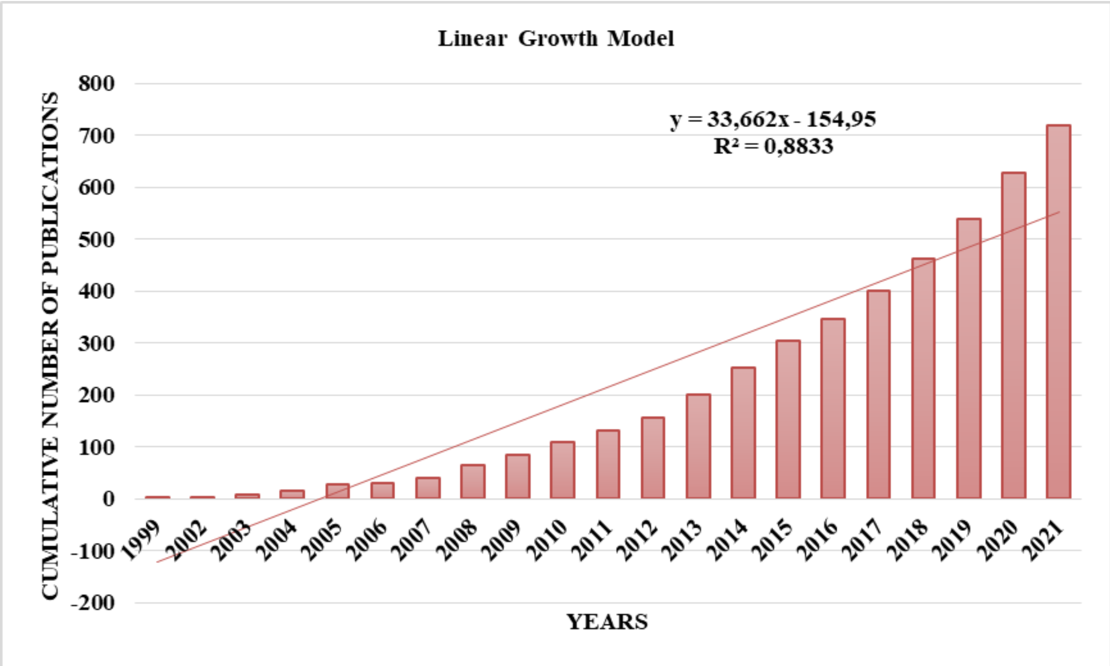


**Figure 2.1(a). Indo-US Collaborated Publications Fit into Linear Growth Model**

**Table 2.1(b) Exponential Growth Model using Chi-Square**

| Sr. no. (x) | Year | Observed no. of publications | Cumulative frequency Y (f) | Expected no. of publications (p) = $3.7157e^{0.283x}$ | (f- p) | $(f-p)^2$ | $(f-p)^2/p$ |
|---|---|---|---|---|---|---|---|
| 1 | 1999 | 1 | 1 | 4.93 | -3.93 | 15.45 | 3.13 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 2 | 2002 | 1 | 2 | 6.54 | -4.54 | 20.65 | 3.16 |
| 3 | 2003 | 6 | 8 | 8.68 | -0.68 | 0.47 | 0.05 |
| 4 | 2004 | 7 | 15 | 11.52 | 3.48 | 12.08 | 1.05 |
| 5 | 2005 | 12 | 27 | 15.29 | 11.71 | 137.04 | 8.96 |
| 6 | 2006 | 3 | 30 | 20.30 | 9.70 | 94.18 | 4.64 |
| 7 | 2007 | 10 | 40 | 26.93 | 13.07 | 170.74 | 6.34 |
| 8 | 2008 | 25 | 65 | 35.74 | 29.26 | 856.02 | 23.95 |
| 9 | 2009 | 19 | 84 | 47.43 | 36.57 | 1337.20 | 28.19 |
| 10 | 2010 | 26 | 110 | 62.95 | 47.05 | 2214.10 | 35.17 |
| 11 | 2011 | 21 | 131 | 83.53 | 47.47 | 2253.10 | 26.97 |
| 12 | 2012 | 25 | 156 | 110.85 | 45.15 | 2038.15 | 18.39 |
| 13 | 2013 | 44 | 200 | 147.11 | 52.89 | 2797.27 | 19.01 |
| 14 | 2014 | 53 | 253 | 195.23 | 57.77 | 3337.87 | 17.10 |
| 15 | 2015 | 51 | 304 | 259.08 | 44.92 | 2018.04 | 7.79 |
| 16 | 2016 | 42 | 346 | 343.81 | 2.19 | 4.78 | 0.01 |
| 17 | 2017 | 55 | 401 | 456.26 | -55.26 | 3053.94 | 6.69 |
| 18 | 2018 | 62 | 463 | 605.49 | -142.49 | 20303.55 | 33.53 |
| 19 | 2019 | 76 | 539 | 803.53 | -264.53 | 69974.02 | 87.08 |
| 20 | 2020 | 89 | 628 | 1066.33 | -438.33 | 192135.16 | 180.18 |
| 21 | 2021 | 91 | 719 | 1415.09 | -696.09 | 484546.17 | 342.41 |
| | **Grand Total** | **719** | **4522** | | | | **853.83** |

The above table depicts the application of the exponential growth model on the bioinformatics literature produced as a result of Indo-US collaboration during the time period under study using the chi-square goodness of fit test. The chi-square value has been calculated and then compared with the theoretical chi-square value at df equals to 20 at 0.05 level of significance.

Calculated chi-square value = **853.83**

Theoretical chi-square value for df (20) at 0.05 level of significance = **31.41**

Exponential growth equation **($y = 3.7157e^{0.283x}$)** and **$R^2 = 0.8961$** has been computed with the help of MS-Excel statistical operations. *The calculated chi-square value is much greater than the theoretical value, so the literature does not follow the exponential growth model.*

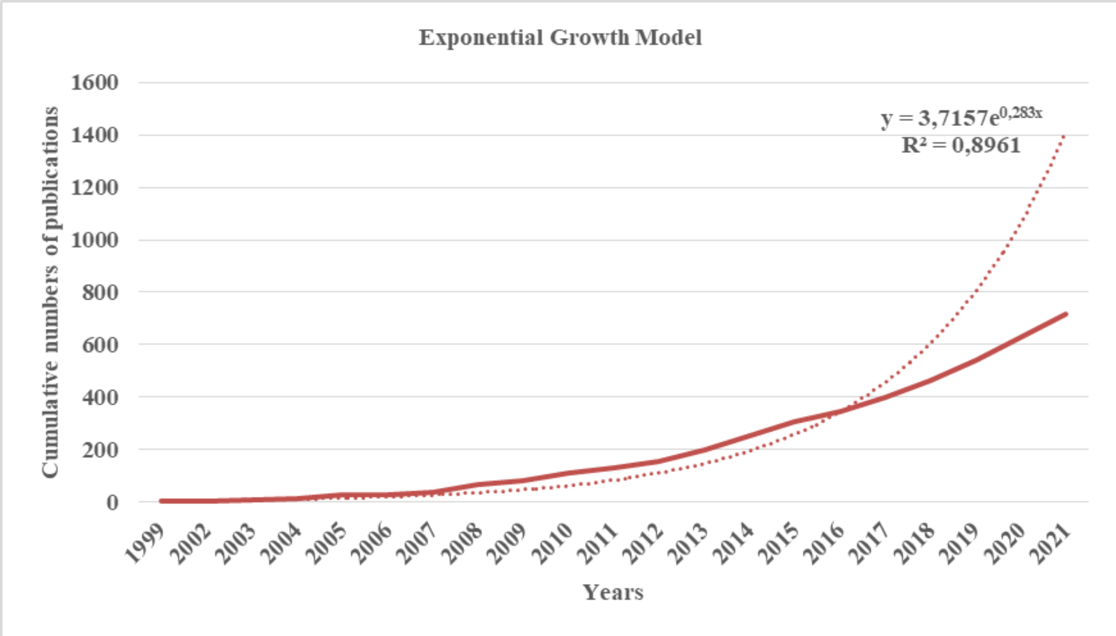


**Figure 2.1(b). Indo-US Collaborated Publications Fit into Exponential Growth Model**

**Table 2.1(c) Power Growth Model using Chi-Square**

| Sr. no. (x) | Year | Observed no. of papers | Cumulative frequency Y (f) | Expected no. of publications (p) = $0.6338x^{2.2589}$ | (f- p) | $(f-p)^2$ | $(f-p)^2/p$ |
|---|---|---|---|---|---|---|---|
| 1 | 1999 | 1 | 1 | 0.63 | 0.37 | 0.13 | 0.21 |
| 2 | 2002 | 1 | 2 | 3.03 | -1.03 | 1.07 | 0.35 |
| 3 | 2003 | 6 | 8 | 7.58 | 0.42 | 0.18 | 0.02 |
| 4 | 2004 | 7 | 15 | 14.52 | 0.48 | 0.23 | 0.02 |
| 5 | 2005 | 12 | 27 | 24.04 | 2.96 | 8.79 | 0.37 |
| 6 | 2006 | 3 | 30 | 36.28 | -6.28 | 39.49 | 1.09 |
| 7 | 2007 | 10 | 40 | 51.40 | -11.40 | 129.91 | 2.53 |
| 8 | 2008 | 25 | 65 | 69.49 | -4.49 | 20.19 | 0.29 |
| 9 | 2009 | 19 | 84 | 90.68 | -6.68 | 44.56 | 0.49 |
| 10 | 2010 | 26 | 110 | 115.04 | -5.04 | 25.41 | 0.22 |
| 11 | 2011 | 21 | 131 | 142.68 | -11.68 | 136.35 | 0.96 |
| 12 | 2012 | 25 | 156 | 173.67 | -17.67 | 312.09 | 1.80 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 13 | 2013 | 44 | 200 | 208.08 | -8.08 | 65.35 | 0.31 |
| 14 | 2014 | 53 | 253 | 246.00 | 7.00 | 48.95 | 0.20 |
| 15 | 2015 | 51 | 304 | 287.49 | 16.51 | 272.54 | 0.95 |
| 16 | 2016 | 42 | 346 | 332.61 | 13.39 | 179.22 | 0.54 |
| 17 | 2017 | 55 | 401 | 381.43 | 19.57 | 383.04 | 1.00 |
| 18 | 2018 | 62 | 463 | 434.00 | 29.00 | 841.14 | 1.94 |
| 19 | 2019 | 76 | 539 | 490.38 | 48.62 | 2364.34 | 4.82 |
| 20 | 2020 | 89 | 628 | 550.62 | 77.38 | 5988.27 | 10.88 |
| 21 | 2021 | 91 | 719 | 614.77 | 104.23 | 10863.68 | 17.67 |
| | **Grand Total** | **719** | **4522** | | | | **46.65** |

The above table depicts the power growth model applied to the literature published in bioinformatics as a result of Indo-US collaboration using the chi-square goodness of fit test. The power growth equation, i.e., $\mathbf{y = 0.6338x^{2.2589}}$ and $\mathbf{R^2 = 0.9914}$**,** has been computed using MS Excel statistical operations. The chi-square value has been calculated and then compared with the theoretical chi-square value at df equals to 20 at 0.05 level of significance.

Calculated chi-square value **= 46.65**
Theoretical chi-square value for df (20) at 0.05 level of significance **= 31.41**
*The calculated chi-square value is much greater than the theoretical value, so the literature does not follow the power growth model.*

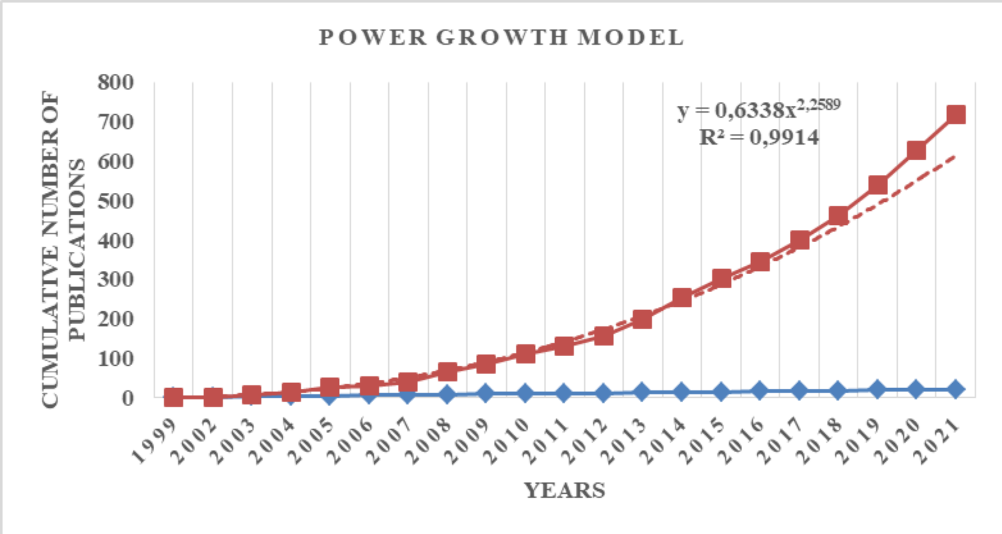


**Figure 2.1(c). Indo-US Collaborated Publications Fit into Power Growth Model**

**Table 2.1(d). Logarithmic Growth Model using Chi-Square**

| Sr. no. (x) | Year | Observed no. of papers | Cumulative frequency Y (f) | Expected no. of publications (p) = 209.73ln(x)-237.88 | (f- p) | $(f-p)^2$ | $(f-p)^2/p$ |
|---|---|---|---|---|---|---|---|
| 1 | 1999 | 1 | 1 | -237.88 | 238.88 | 57063.65 | -239.88 |
| 2 | 2002 | 1 | 2 | -92.51 | 94.51 | 8931.43 | -96.55 |
| 3 | 2003 | 6 | 8 | -7.47 | 15.47 | 239.26 | -32.04 |
| 4 | 2004 | 7 | 15 | 52.87 | -37.87 | 1433.95 | 27.12 |
| 5 | 2005 | 12 | 27 | 99.67 | -72.67 | 5280.55 | 52.98 |
| 6 | 2006 | 3 | 30 | 137.91 | -107.91 | 11643.64 | 84.43 |
| 7 | 2007 | 10 | 40 | 170.24 | -130.24 | 16961.35 | 99.63 |
| 8 | 2008 | 25 | 65 | 198.24 | -133.24 | 17753.24 | 89.55 |
| 9 | 2009 | 19 | 84 | 222.94 | -138.94 | 19305.41 | 86.59 |
| 10 | 2010 | 26 | 110 | 245.04 | -135.04 | 18236.12 | 74.42 |
| 11 | 2011 | 21 | 131 | 265.03 | -134.03 | 17964.20 | 67.78 |
| 12 | 2012 | 25 | 156 | 283.28 | -127.28 | 16200.06 | 57.19 |
| 13 | 2013 | 44 | 200 | 300.07 | -100.07 | 10013.37 | 33.37 |
| 14 | 2014 | 53 | 253 | 315.61 | -62.61 | 3919.95 | 12.42 |
| 15 | 2015 | 51 | 304 | 330.08 | -26.08 | 680.13 | 2.06 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 16 | 2016 | 42 | 346 | 343.62 | 2.38 | 5.69 | 0.02 |
| 17 | 2017 | 55 | 401 | 356.33 | 44.67 | 1995.42 | 5.60 |
| 18 | 2018 | 62 | 463 | 368.32 | 94.68 | 8964.74 | 24.34 |
| 19 | 2019 | 76 | 539 | 379.66 | 159.34 | 25390.13 | 66.88 |
| 20 | 2020 | 89 | 628 | 390.41 | 237.59 | 56446.67 | 144.58 |
| 21 | 2021 | 91 | 719 | 400.65 | 318.35 | 101348.19 | 252.96 |
| | **Grand Total** | **719** | **4522** | | | | **813.46** |

The above table depicts the application of the logarithmic growth model on the literature published in bioinformatics as a result of Indo-US collaboration during the period under study using the chi-square goodness of fit test. The chi-square value has been calculated and then compared with the theoretical chi-square value at df equals to 20 at 0.05 level of significance.

The logarithmic growth equation, **i.e., $y = 209.73\ln(x) - 237.88$** and **$R^2 = 0.5953$,** has been calculated using MS-Excel statistical operations.

Calculated chi-square value **= 813.46**
Theoretical chi-square value for df (20) at 0.05 level of significance **= 31.41**
*The calculated chi-square value is much greater than the theoretical value, so the literature does not follow the logarithmic growth model.*

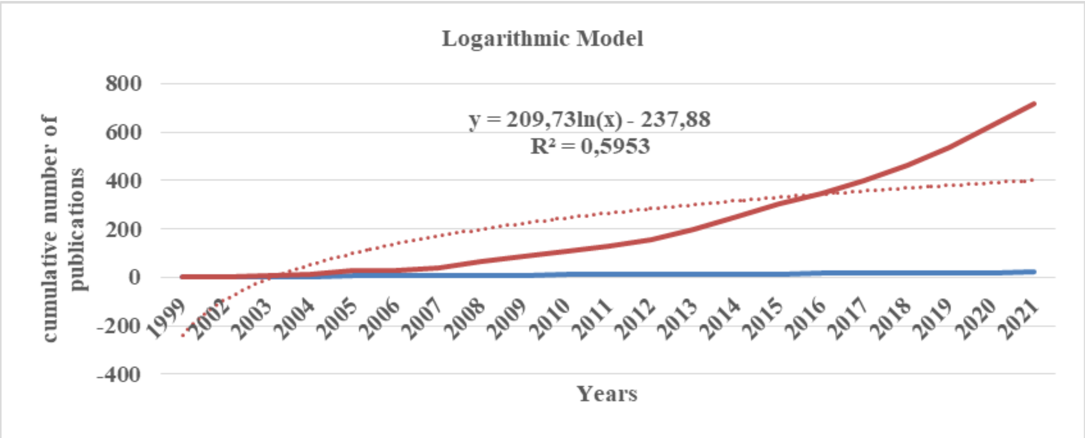


**Figure 2.1(d). Indo-US Collaborated Publications Fit into Logarithmic Growth Model**

### 2.2 APPLICATION OF GROWTH MODELS USING CURVE FIT VALUES

**Table 2.2. $R^2$, F, Y Values of Different Growth Models**

| Name of Growth model | | $R^2$ | F value | Y |
|---|---|---|---|---|
| **1.** | **Linear Growth Model** | 0.8833 | 1.132182 | 33.662x - 154.95 |
| **2.** | **Exponential Growth Model** | 0.8961 | 0.321579 | $3.7157e^{0.283x}$ |
| **3.** | **Power Growth Model** | 0.9914 | 1.292789 | $0.6338x^{2.2589}$ |
| **4.** | **Logarithmic Growth Model** | 0.5953 | 1.679804 | 209.73ln(x) - 237.88 |

The above table depicts the Coefficient of Regression, i.e., $R^2$ values, the F statistics of different growth models, and the Y, i.e., the equation of various growth models. The percentage of the variance between the observed values of the data under study and the fitted values of the data is the $R^2$. If the value of the coefficient of regression is higher for data, that means there is less difference between the observed and expected values. And, if the $R^2$ value is less for any data, then a large difference is seen between the observed and the expected values. Here, the value of the coefficient of regression ($R^2$) is high, which ensures less difference between the observed and the expected values. The F values and the equations of different growth models based on the present dataset have been computed using MS Excel software.

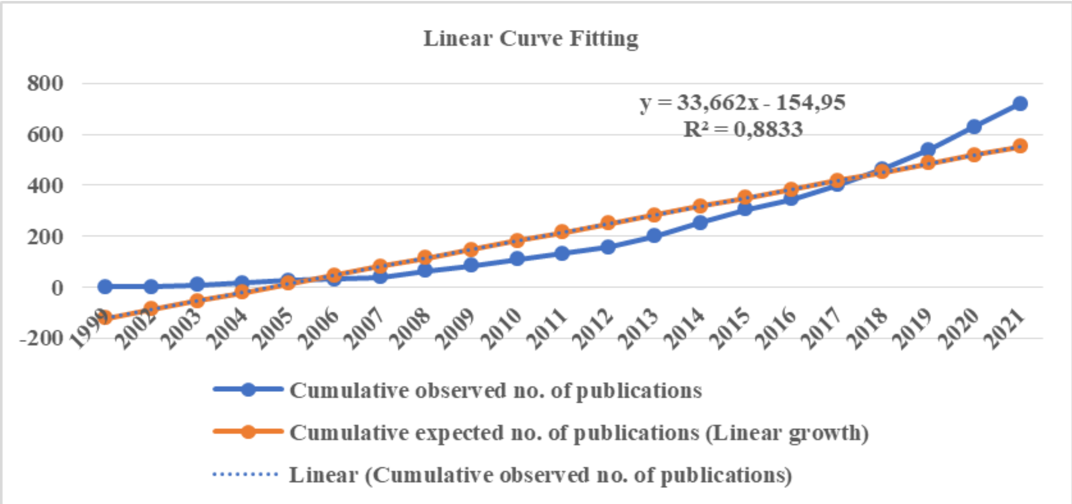

**Figure 2.2(a). Linear Curve Fitting of Indo-US Collaborated Publications**

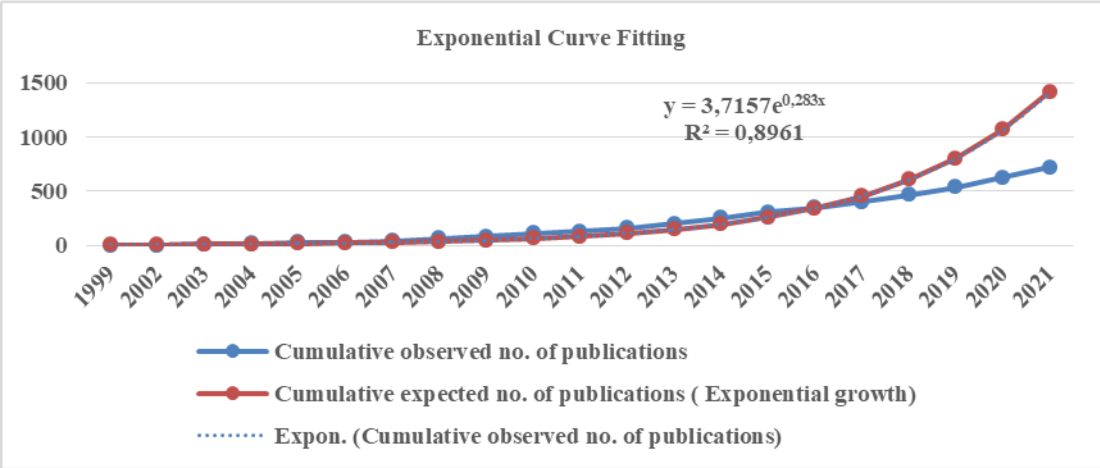


**Figure 2.2 (b). Exponential Curve Fitting of Indo-US Collaborated Publications**

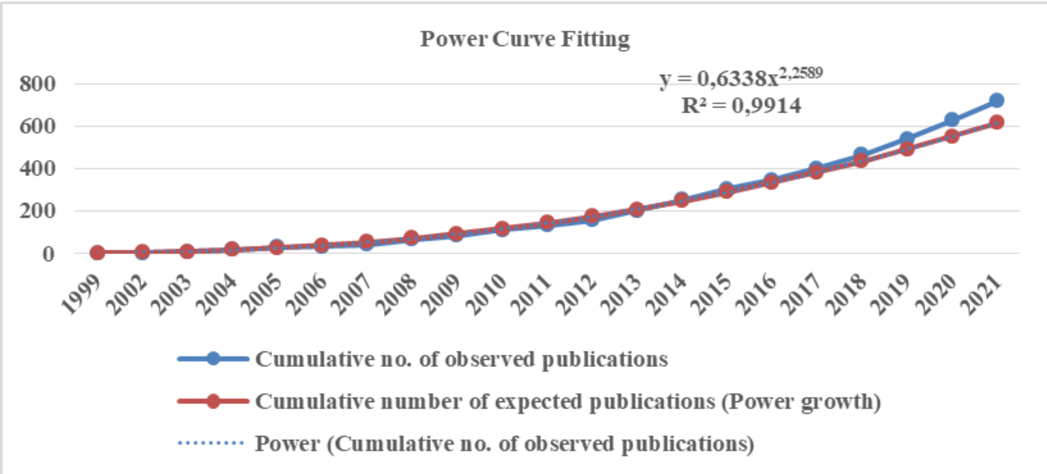


**Figure 2.2 (c). Power Curve Fitting of Indo-US Collaborated Publications**

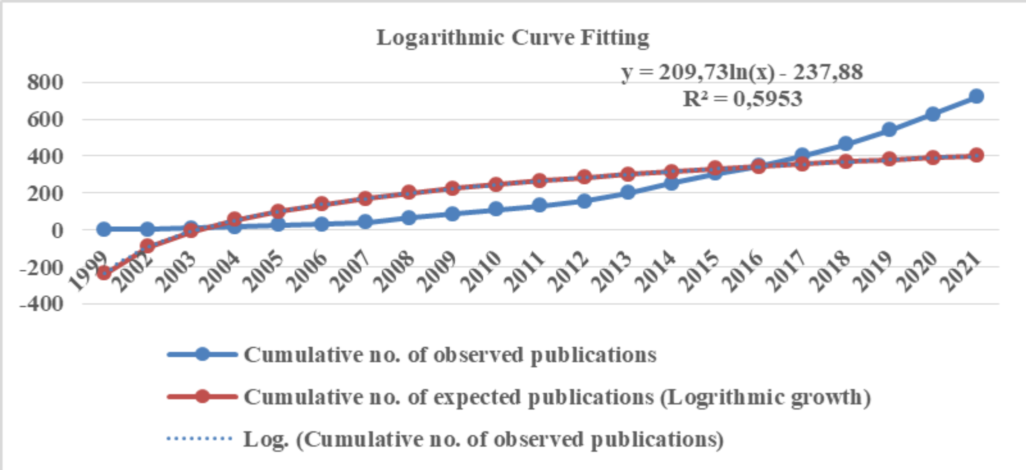


**Figure 2.2(d). Logarithmic Curve Fitting of Indo-US Collaborated Publications**

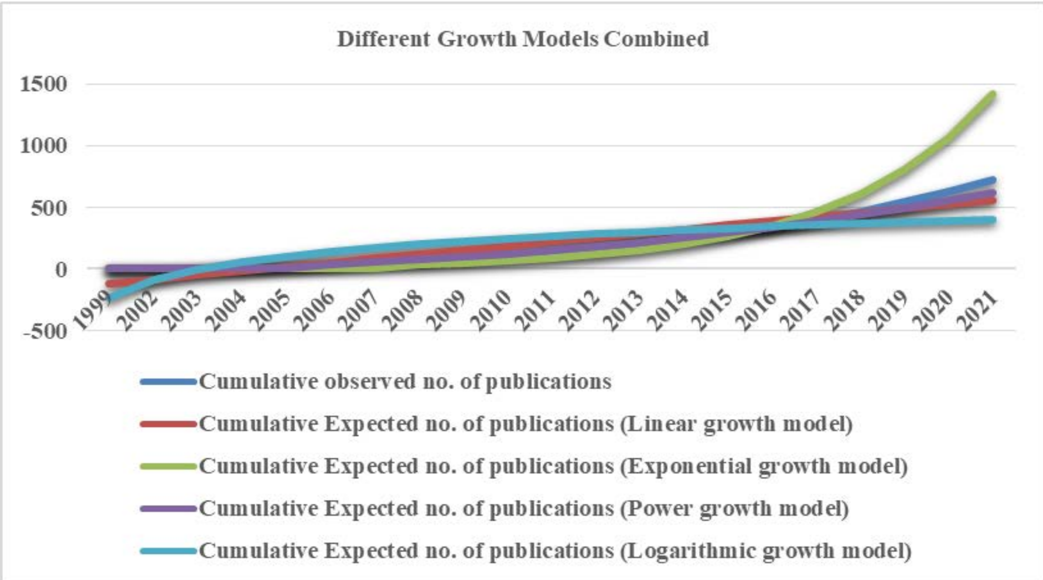


**Figure 2.2(e). Different Growth Models Combined**

The above figures 2.2(a), 2.2(b), 2.2(c), 2.2(d), and 2.2(e) graphically represent the application of different growth models using curve fit values. The cumulative number of observed publications is plotted along with the cumulative number of expected publications during the time period under study.

The cumulative expected number of publications has been calculated using different operations related to growth analysis, i.e., using the growth model equations that are formed as a result of different growth model analyses in MS Excel software. The close relationship of the curves has been observed as a result of growth model analysis based on observed and expected no. of publications. With an $R^2$ value of 0.9914 (99% variance), the power growth curve was found to be the best match or the best fitting curve for the literature on bioinformatics produced as a result of collaboration between India and the USA, followed by the exponential growth curve having $R^2$ value equivalent to 0.8961 (89% variance), linear growth curve having $R^2$ value equal to 0.8833 (88% variance) and logarithmic growth curve having the $R^2$ value equal to 0.5953 (60% variance).

## FINDINGS

**1.** The years with the most and lowest number of publications are 2021 (91 articles, 12.66%) and 1999 and 2002 (1 paper each, 0.14%). The average AGR, RGR, and DT values, respectively, are 53.49, 0.33, and 3.30. The average annual growth rate (AGR) is 53.49, indicating an overall upward trend in the growth of publications in bioinformatics. Furthermore, the declining RGR tends to reflect the slow but steady increase in the number of articles published, as evidenced by an increase in DT, which is significantly greater than RGR during the course of the study *(Table 1 ).*

The M-K test for trend observation shows that the z statistics value is positive and greater than the z critical value; hence, a positive or increasing trend occurs in the dataset and the trend is significant at 99% confidence level *(Table 1.1).*

**Hypothesis I** stating that "*There is an overall increasing trend in literature produced as a result of collaboration from India and the USA in the bioinformatics domain over the period of time*" stands supported.

**2.** The different growth models have been applied to the present dataset using the chi-square test, and it is found that the calculated chi-square value, i.e., -34.473, is less than the theoretical value of chi-square at df = 20, i.e., 31.41 when calculated using linear growth model. *Hence, the growth of literature produced as a result of Indo-US partnership in the area of bioinformatics follows the linear growth model.*

**The Hypothesis II** stating that "*There is a linear growth of literature in bioinformatics produced as a result of Indo-US collaboration*" stands supported.

**3.** The different growth models using curve fit values have been applied, and it is found that with an $R^2$ value of 0.9914 (99% variance), the power growth curve was found to be the best match, followed by an exponential growth curve having an $R^2$ value equivalent to 0.8961 (89% variance), linear growth curve having $R^2$ value equal to 0.8833 (88% variance), and logarithmic growth curve having the $R^2$ value equal to 0.5953 (60% variance) *(Table 2.2).*

**CONCLUSION:** Bioinformatics is a relatively new field of science in which research activity is steadily expanding. This study will be beneficial to data managers, researchers, teachers, and academicians working in the field of bioinformatics in terms of understanding the current state of collaborative work between India and the USA in bioinformatics in the form of literature and how it is growing over the years. Overall, an increasing trend has been observed, and the declining RGR tends to reflect the slow but steady increase in the number of articles published, as evidenced by an increase in DT. The growth of literature produced as a result of Indo-US partnership in the area of bioinformatics follows the linear growth model. When the different growth models using curve fit values based on $R^2$ have been applied, the power growth curve is found to be the best match, followed by the exponential growth curve, linear growth curve and logarithmic growth curve.

## REFERENCES

Baskaran, C. (2016). A Scientometric study on bioinformatics literature during 1999-2013. *International Journal of Library Science and Information Management*, 2(4). https://www.researchgate.net/publication/324537363

Baskaran, S. (2021). Scientometric study of bioinformatics research based on *Web of Science* database during 2010-2019. *International Journal of Research in Library Science*, *7*(1), 196–205. https://doi.org/10.26761/ijrls.7.1.2021.1389

Elangovan, N. (2020). Application of growth models to research literature of all India Institute of Medical Sciences. *DESIDOC Journal of Library & Information Technology, 40*(03), 153–159. https://doi.org/10.14429/djlit.40.03.14876

El-Shorbagy, M., & El-Refaey, A. M. (2022). COVID-19: Mathematical growth vs. precautionary measures in China, KSA, and the USA. *Informatics in Medicine Unlocked, 28*, 100834. https://doi.org/10.1016/j.imu.2021.100834

Hadagali, G. & Anandhalli, G. (2015). Modeling the growth of neurology literature. *Journal of Information Science Theory and Practice*. 3. 45-63. 10.1633/JISTaP.2015.3.3.3.

Jozi, Z., & Nourmohammadi, H. (2022). Scientometrics analysis of world scientific research of pathology and forensic medicine. *Iranian Journal of Pathology*, 191–201. https://doi.org/10.30699/ijp.2022.541660.2756

Kumar, R.S. & Kaliyaperumal, K. (2015). A scientometric analysis of mobile technology publications, *Scientometrics*, 105(2), 921-939. DOI: 10.1007/s11192-015-1710-7

Mahapatra, M. (1985). On the validity of the theory of exponential growth of scientific literature. In *Proceedings of the 15th IASLIC Conference*, Bangalore, 61-70.

Manoharan, A., Kanagavel, B., Muthuchidambaram, A., & Kumaravel, J. P. S. (2011). Bioinformatics research: An informetric view. *IPCSIT*, *16*, 199-204.

Nayak, S. N., & Bankapur, V. M. (2017). Modelling the growth of global agricultural literature: A scienometric study based on CAB-Abstracts. International Journal of Library and Information Studies, 7(3). http://www.ijlis.org

Neelamma, G. & Anandhalli, G. (2020). Modelling the growth of literature in the area of crystallography. *Library Philosophy and Practice (e-journal)*. 3813. https://digitalcommons.unl.edu/libphilprac/3813

Rao, I. K. R. (2010). *Growth of literature and measures of scientific productivity: Scientometric models*. Ess Ess Publications.

Sangam, S.L., Madalli, D. & Arali, U.B. (2015). Scientometrics profile of global genetics literature as seen through *PubMed*. *COLLNET Journal of Scientometrics and Information Management,* 9(2), 175-192. DOI: 10.1080/09737766.2015.1069956

Sharma, P., Gupta, B.M. & Kumar, S. (2002). Application of growth models to science and technology literature in research specialties. *DESIDOC Bulletin of Information Technology*. 22. 10.14429/dbit.22.2.3560.

Shilpa, B., & Padmamma, S. (2020). Scientometrics observations of authorship pattern and collaborative research on leukemia. *International Journal of Information Dissemination and Technology*, 10(1), 52. https://doi.org/10.5958/2249-5576.2020.00008.4

Singh, K. P., & Verma, S. (2021). Modeling the growth of food science and technology literature in India. *Library Philosophy and Practice,* 1-12. https://www.proquest.com/scholarly-journals/modeling-growth-food-science-technology/docview/2492708613/se-2